Compute Working Team Update

State of the Project

Charles Doutriaux, Dan Duffy, Jason Boutte and
Thomas Maxwell, …

December 6th, 2017

LLNL-PRES-742786
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Outline

- Take Home
- Accomplishments
- Things to Come
- Resources

Take Home

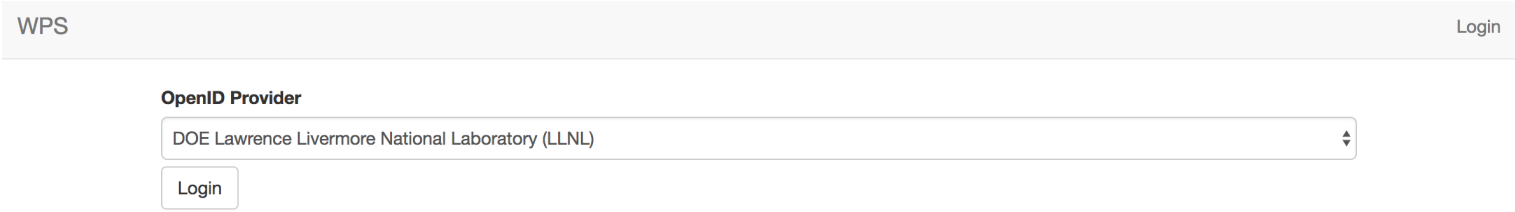
- Many Backends Matured
- All compatible via End User API
- COG Integrated Front End
- Ready to be considered as part of installation

Server Side Accomplishments: API

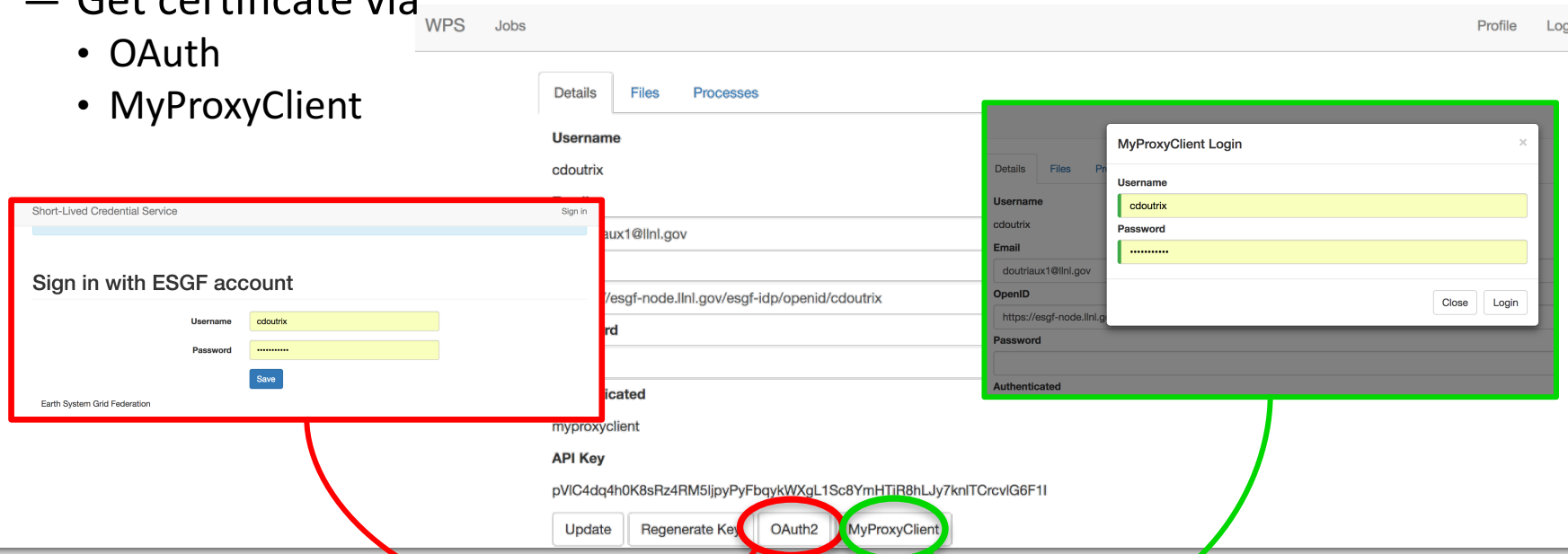
- No changes to Server API!

Server Side Accomplishments: Security

- Authentication added
  - Use OpenID to login

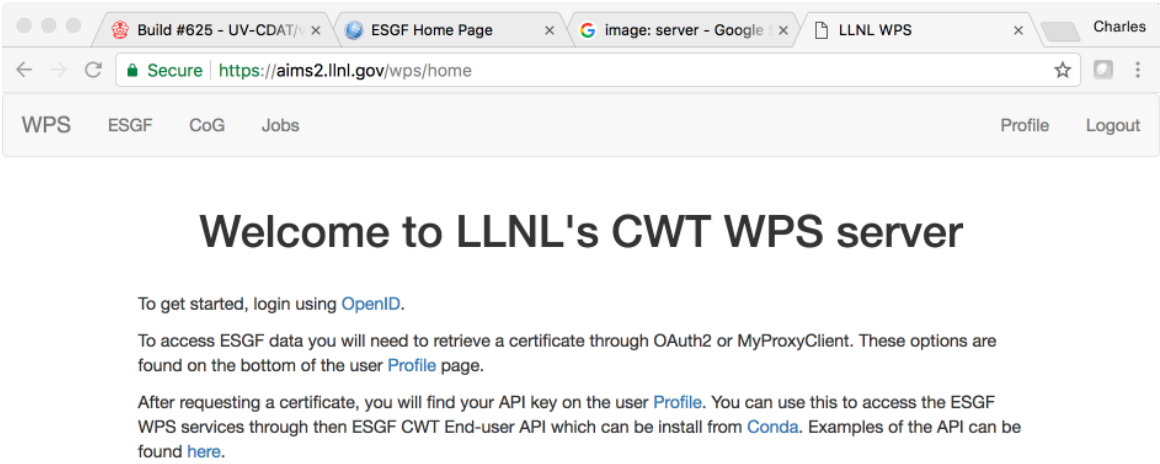


  - Get certificate via
    - OAuth
    - MyProxyClient



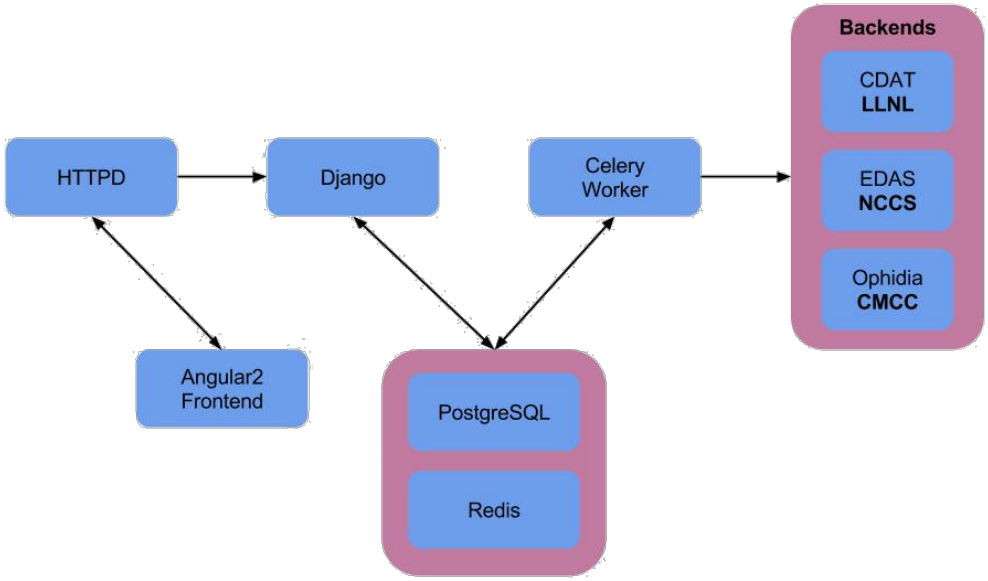
Server Side Accomplishments: Server

- Official Server Implemented:
  - Code: https://github.com/ESGF/esgf-compute-wps
  - Docker-based
  - Re-engineered core
  - Celery queues to manage submitted processes
  - Includes LLNL, NASA/EDAS and Ophidia's services
- Deployed at LLNL: https://aims2.llnl.gov/wps/home



Server Side Accomplishments: LLNL

- Services are CDAT-based
  - Aggregation
  - Average
  - Regrid
  - Min/Max



Server Side Accomplishments: NASA

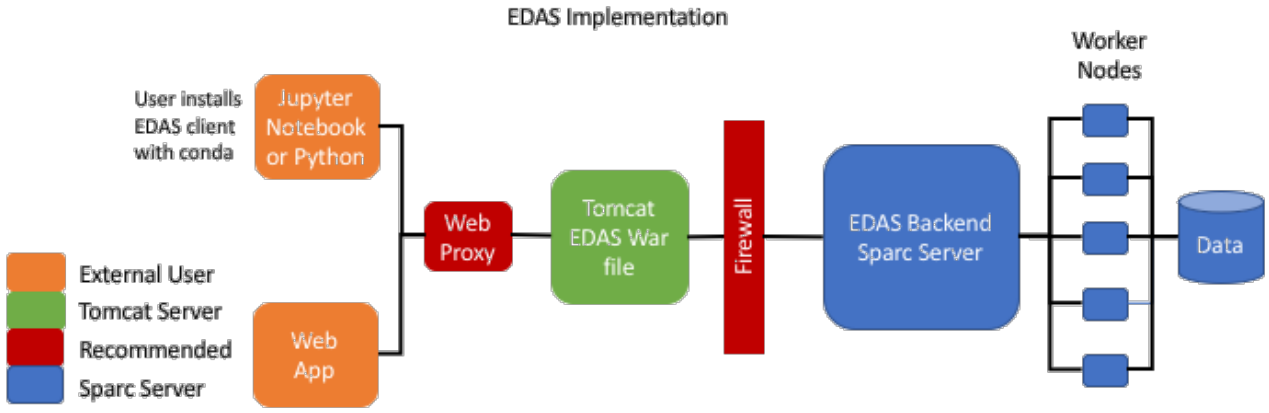
- NASA EDAS v. 1.1 deployment on DASS goes public:
  - *https://edas.nccs.nasa.gov/wps/cwt*

- Current Open Source EDAS Distribution:
  - Server: *https://github.com/nasa-nccs-cds/EDAS.git*
  - Web app: *https://github.com/nasa-nccs-cds/CDWPS.git*
  - Client: *https://github.com/ESGF/esgf-compute-api.git*

- Documentation and sample Jupyter Notebooks available at:
  - *https://www.nccs.nasa.gov/services/Analytics*

- List available canonical operations (kernels):
  - *https://edas.nccs.nasa.gov/wps/cwt?request=GetCapabilities*
  - *13 available: emul, ediff, min, emin, max, emax, sum, esum, avg, eavg, rms, erms, ediv*

- List available data collections (17 available: MERRA, etc…):
  - *https://edas.nccs.nasa.gov/wps/cwt?request=GetCapabilities&identifier=coll*

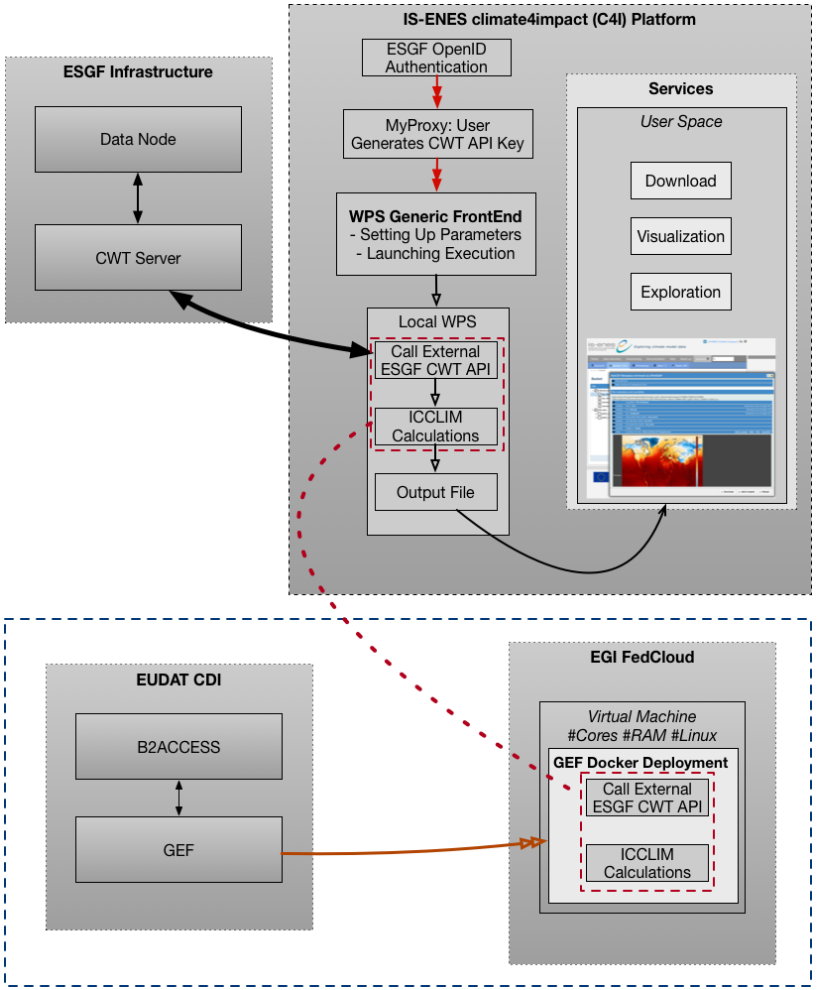


Server Side Accomplishments: Ophidia

- ESGF CWT Plugin for Ophidia implemented
- Properly translates WPS requests into Ophidia requests
- Based on PyOphidia (available on Conda)
- Leverages the workflow support of Ophidia (through the PyOphidia wsubmit) to implement a WPS process
- Flexible mechanism since each WPS process can be based on several Ophidia operators (workflow)
- Available functionalities: subsetting along any dimension (space and time), maximum & minimum along a specific dimension
- Deployment of the CWT module in the OphidiaLab environment at CMCC

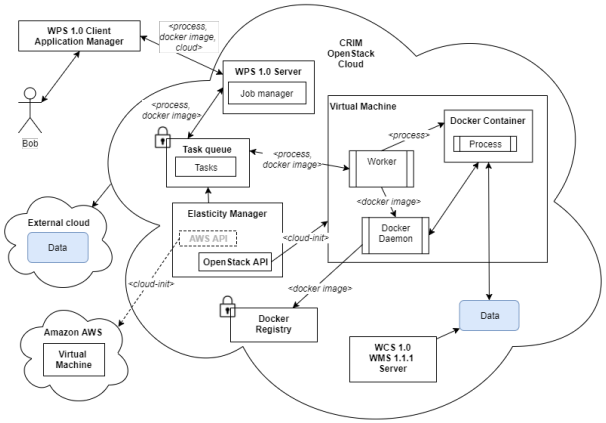
Server Side Accomplishments: CERFACS

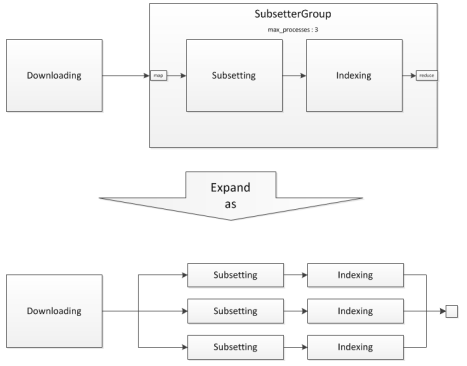
- Interfaced CWT and the IS-ENES Climate4impact platform
- Access CWT WPS from within a EUDAT-GEF Execution on EGI



Server Side Accomplishments: CRIM

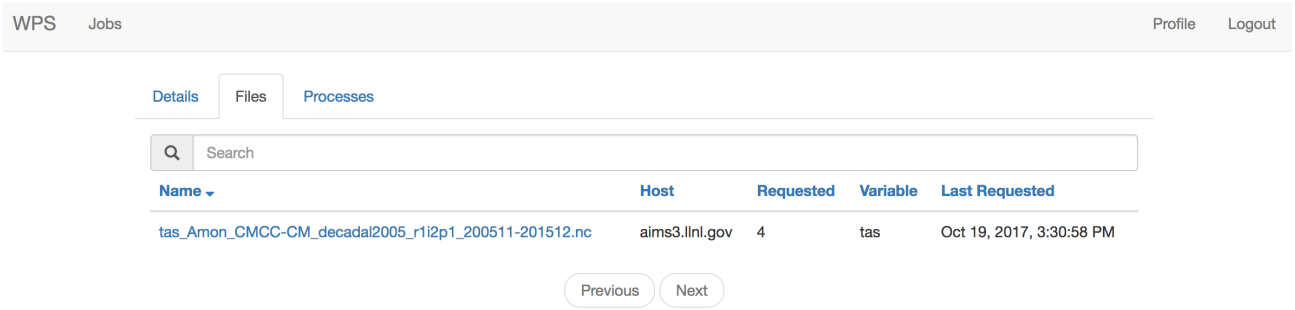
- Advanced parallel workflow execution
- Prototyped an authorization mechanism for data and processes
- Advanced WPS hybrid cloud execution to integrate with CWT API
- Communicated ESGF uses cases to OGC (security, server-side API, infrastructure, etc.)
- Started work on common test suite
- Contributed to OpenClimateGIS and Birdhouse
- For 2018: more work required to integrate, transfer and harmonize PAVICS and OGC advancements into ESGF software



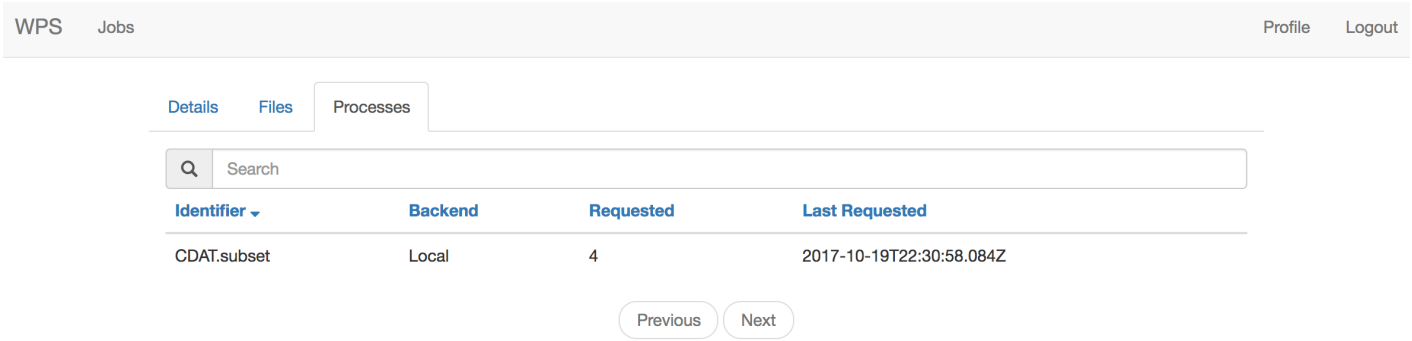


Server Side Accomplishments: Analytics

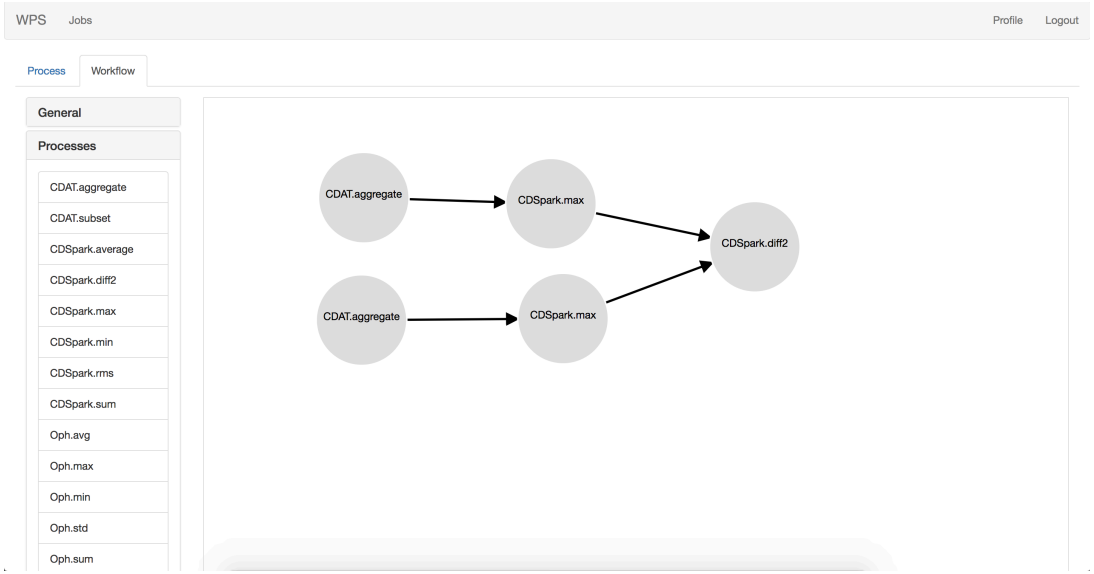
- We are already capturing:
  - Files accessed

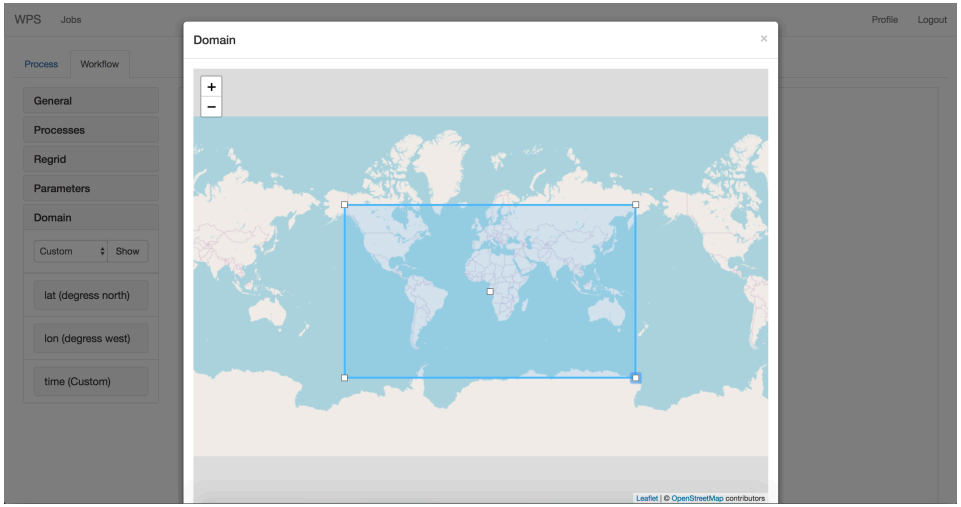


  - Process launched

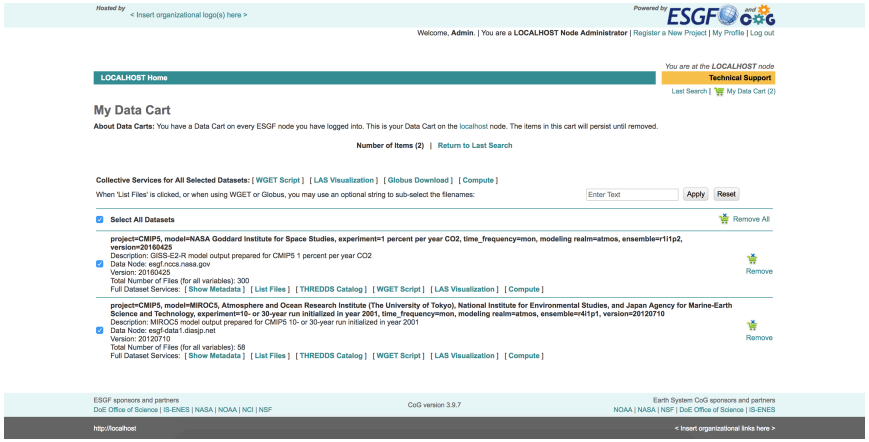


Work In Progress: Workflow and Provenance

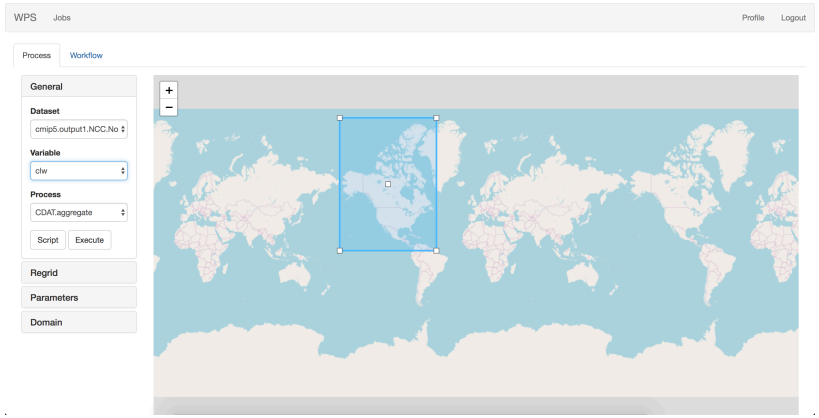


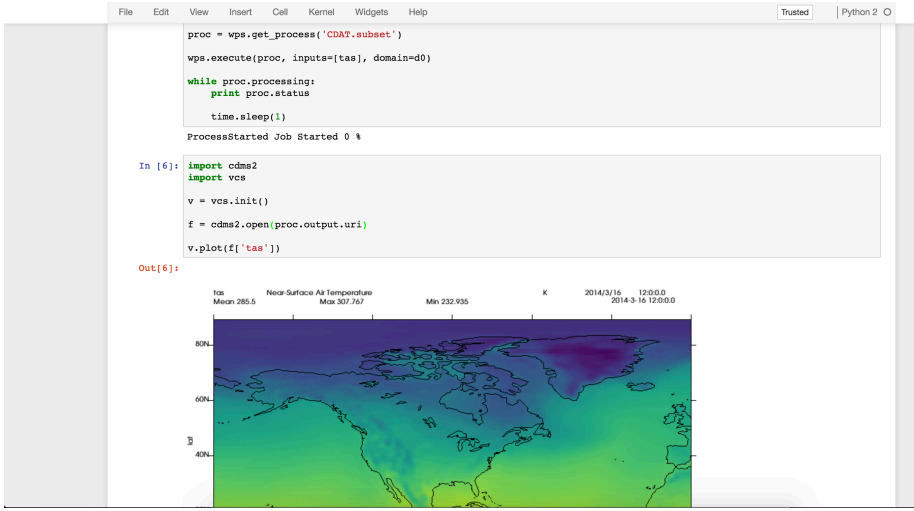


COG Integration



https://aims2.llnl.gov





Things to come

- Full support for OAuth
- Integrated in ESGF release cycle
- Workflows finalized
- Helping other teams' work to be compatible with end-user API
  - Ouranos/Pavics
- Documenting Services
- More Advanced Caching
- Fully distributed
  - Scalability
  - Discovery
- More Services

Resources

- Email: esgf-cwt@llnl.gov
- Webex:
  - First Monday of the month: General Meeting
  - Third Monday of the month: Implementation Meeting
- Documentation
  - Web(requires login)
  - API
- Code (github)
  - Server: https://github.com/ESGF/esgf-compute-wps
  - End-user: https://github.com/ESGF/esgf-compute-apiResources

Questions Suggested on Agenda

- Define a scalable compute resource (clusters and HPCs) for ESGF data analysis
- Data analytical and visualization capabilities and services
- Performance of model execution
- Advanced networks as easy-to-use community resources (i.e., resource management)
- Provenance and workflow
- Automation of steps for the computational work environment
- Resource management, installation and customer support
- Identify key gaps, identify benefitting communities, and prioritize next steps
- Analysis services when multiple data sets are not co-located (future work)